NASA TM X-1434

METRIZABILITY AND PATTERN RECOGNITION

By Peter Duran

Electronics Research Center
Cambridge, Mass.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

METRIZABILITY AND PATTERN RECOGNITION

By Peter Duran
Electronics Research Center

SUMMARY

Some attempts have been made to apply metric space theory to pattern recognition. The relationship of metrizability to pattern recognition problems will be investigated. The number of elements (cardinality) in the set of patterns also has a strong bearing on the ability to use metric concepts in classification of the patterns. The relationship of cardinality to metrizability is discussed.

INTRODUCTION

One of the major tasks in pattern recognition is the classification of those things which are equivalent with respect to a given criterion. It is sometimes convenient to introduce a function on the set of data points which will, in some sense, indicate how the data points appear to cluster. Such a function would resemble the ordinary distance function that is used in metric space theory. The set of data points would be partitioned into disjoint, non-empty classes which would exhaust the set if some appropriate equivalence relation induced by the distance function were used.

If a distance function is associated with the set of data points, the properties of the topology resulting from the metric can be used to classify the data points. For example, we may define the data points x and y to be equivalent if they lie in the same component of the topological space. A component of a topological space is a maximal connected subset of the space. In other words, with respect to the set of data points, a component is a subset of the data points, which is connected and contained in no larger connected subset of the data points. In any proposed equivalence relation on the set of data points it would be required that the limit of a sequence of points of a given equivalence class should belong to that class. Since the components of the topological space are closed, it is appropriate to use them as equivalence classes.

The structure of a topology determined by a metric is also a function of the cardinality of the space. The cardinality of the space will determine in certain instances whether or not the space is metrizable. If the space is metrizable, the cardinality of the space will also determine in some instances the structure of the components.

In some cases it will be seen that the components of the data set are just the individual data points. This means that no two data points are equivalent. Under other circumstances it will be seen that the whole data set is a component. This implies that all of the data points are equivalent. The following theorems describe the usefulness of metrizability concepts in pattern recognition. A further investigation into metrics yielding non-trivial partitions of a data set might profitably be pursued.

ZERO DIMENSIONALITY, METRIZABILITY, AND THEIR RELATION TO CARDINALITY

Many definitions of dimensionality have been given for very general topological spaces. A satisfactory theory of dimension has been given for separable metric spaces. A defining condition for a topology will be given such that the dimension and metrizability of the topological space is a function of the cardinality of the space. A large class of spaces will be shown to be non-metrizable.

Let X be a set such that $1 \leq C(X) < \aleph_0$, where C(X) is the cardinality of X. Since a topology for X is T_1 if, and only if, $\forall\ x \in X$, $\{x\}$ is closed, take as a subbase for a topology on X the family of subsets $\{X - \{y\} \mid y \in X\}$. This subbase generates the smallest, with respect to inclusion, T_1-topology for X. This topology for X is T_2 and discrete, since finite and T_1 if, and only if, finite and T_2 if, and only if, finite and discrete. It is easily shown that this minimal T_1-topology for X is compact, separable totally disconnected, and satisfies the first and second axioms of countability. The topology is metrizable by the following metric: $\forall\ x, y \in X$, if $x = y$, then $d(x, y) = 0$; if $x \neq y$, then $d(x, y) = 1$.

<u>Definitions</u>: A separable metric space X has dimension zero at a point $x \in X$ if, and only if, for all neighborhoods of x, N(x), there is a neighborhood of x, N'(x), such that $N'(x) \subseteq N(x)$ and the boundary of $N'(x) = \emptyset$. A separable metric space has dimension zero if, and only if, it has dimension zero at each of its points.

<u>Remark</u>: A compact, separable metric space is zero dimensional if, and only if, it is totally disconnected.

THEOREM

If X is a non-void finite set with the minimal T_1-topology, then it has dimension zero.

<u>PROOF</u>:

The metrizability and separability of X were indicated above. The assertion follows by virtue of the compactness and total disconnectedness of X. It also follows from the fact that $\forall\ x \in X$, $\{x\}$ is open and closed.|

Let X be a set such that $C(X) \geq \aleph_0$. Again, for a topology, take as a subbase the family of subsets of X, $\{X - \{y\} \mid y \in X\}$. This subbase generates the minimal T_1-topology; the family of all finite intersections of the members of the subbase is a base.

THEOREM

The minimal T_1-topology for a set X where $C(X) = \aleph_0$ satisfies the first and second axioms of countability.

PROOF:

The finite intersections of the members of the subbase correspond to the finite subsets of X, and conversely. The cardinality of the family of finite subsets of a countably infinite set is $\aleph_0$.|

LEMMA

Each non-empty open set in the minimal T_1-topology for a set X where $C(X) \geq \aleph$ contains all but a finite number of the points of X.

PROOF:

Each member of the base is the intersection of a finite number of subbase members; hence, it contains all but a finite number of the points of X. An arbitrary union of members of the base contains each subbase member of the union, and therefore, contains all but a finite number of the points of X.|

THEOREM

The minimal T_1-topology for a set X where $C(X) \geq \aleph_0$ is connected and not T_2.

PROOF:

Assume that $X = 0_1 \cup 0_2$ where 0_1 and 0_2 are non-void open subsets of X and $0_1 \cap 0_2 = \varphi$. $X - 0_1 = 0_2$ and $X - 0_2 = 0_1$. It follows that 0_1 and 0_2 contain all but an infinite number of the points of X. Assume that $x, y \in X$ and $x \neq y$. Let 0_x and 0_y be open subsets of X containing, respectively, x and y. If $0_x \cap 0_y = \varphi$, each of 0_x and 0_y would not contain an infinite number of the points of X.|

COROLLARY

The minimal T_1-topology for a set X where $C(X) \geq \aleph_0$ is not zero dimensional and is not pseudometrizable.

PROOF:

The first assertion follows from the fact that the only subsets of X which are both open and closed are X and $\emptyset$. A space is metrizable if, and only if, it is T_1 and pseudometrizable. X is T_1 and is not metrizable, since it is not T_2; therefore, it is not pseudometrizable.|

THEOREM*

The minimal T_1-topology for a set X where $C(X) \geq \aleph_0$ is compact.

PROOF:

Let $\{0_a\}$ be an open covering of X which is infinite. Select a non-void 0_1 from $\{0_a\}$. $X - 0_1$ is finite and $\{0_a \mid a \neq 1\}$ is an open covering of $X - 0_1$. For $x \in X - 0_1$, select a 0_2 in $\{0_a \mid a \neq 1\}$ which contains x. $\{0_a \mid a \neq 1, 2\}$ is an open covering of $X - (0_1 \cup 0_2)$. Repetition of this process yields a finite subfamily $0_1, 0_2, \ldots, 0_n$ of $\{0_a\}$ which covers X.|

LEMMA

If (X, d) is a metric space such that $2 \leq C(X) < c$, where c is the cardinality of the continuum, then (X, d) is totally disconnected.

PROOF:

Assume that $A \subseteq X$, $C(A) \geq 2$, and A is connected. Define a function f by $f(x) = d(x, a)$ where a is a fixed member of A and $x \in A$. f is continuous. f(A) is connected, as the continuous image of a connected set. $f(a) = 0$. Let $b \in A$ and $a \neq b$, then $f(b) > 0$; hence, f(A) is an interval. It follows that $C(f(A)) = c$ and hence $C(A) \geq c$.|

---

*Theorem is actually a corollary of the lemma on page 3.

COROLLARY

Let $\{(X_i d_i) \mid i \in I\}$ be an indexed set of metric spaces such that $2 \leq C(I) < \aleph_0$ and $\forall i \in I$, $1 \leq C(X_i) \leq \aleph_0$. Then $\Pi\{X_i \mid i \in I\}$ with any metric is totally disconnected.

PROOF:

$C(\Pi\{X_i \mid i \in I\}) \leq \aleph_0.|$

Let (X, d) be a pseudo-metric space such that $C(X) < c$. Define an equivalence relation $\sim$ on X by $\forall x, y \in X$, $x \sim y$ if, and only if, $d(x,y) = 0$. Denote the family of equivalence classes of $\sim$ by X' and denote by x' the equivalence class to which $x \in X$ belongs. Define a function d' on the Cartesian product of X' with itself by $d'(x',y') = d(a,b)$ where $a \in x'$ and $b \in y'$. (X',d') is a metric space.

COROLLARY

(X', d') is totally disconnected.

PROOF:

$C(X') \leq C(X).|$

NON-METRIZABILITY THEOREM

Every topological space the cardinality of which is less than that of the continuum and which contains a connected subset of two or more distinct points is not metrizable.

PROOF:

If the space were metrizable, then it would be totally disconnected.|

COROLLARY

Every T_1-topological space whose cardinality is less than that of the continuum and which contains a connected subset of two or more distinct points is non-pseudo-metrizable.

NASA Electronics Research Center, Cambridge, Massachusetts, August 1967, 125-03-02-27.